Assignment 2 – ECO375

Luis Rojas S.

1003650676

Abstract

Economists Ian Ayres and John Donohue III claim in their famous paper "Shooting Down the More Guns, Less Crime Hypothesis" that there is not enough statistically significance evidence to claim that those states that adopt laws that allow their citizens to carry legally a gun do not help reduce their crime rates, and thus debating the hypothesis claimed by pro-gun economist John Lott. Using the panel data that economists Ayres and Donohue extended to 1999, from the previous research done by Lott only up to 1992, we analyze the hypothesis that if those states that adopted the shall laws in fact reduced their crime. In order to do so, we use a set of panel data methods that include a set of covariates, controlling for fixed and time effects in order to find a causal relationship between the shall-issue laws and three different crimes rates. We found that for the years 1977 to 1999 the issue of these laws did not change, at any statistical significance level, the rates of violent crimes, robbery and murder.

Introduction

In his famous book "*More guns, less crime: Understanding crime and gun-control laws"*, economist John R. Lott states that the number of guns in the population is negatively correlated with the amount of crime in this society. That is if one state may adopt one law that allows residents to carry guns legally, after satisfying some conditions, crime should decrease since delinquents now face a higher probability of failure in their intentions of robbery. These types of laws, called shall-issue law, may be adopted independently by state in the USA. We use a panel data sample between the years 1977 to 1999 where some states started to adopt these laws in different years and others did adopt them at all. The aim is to check if those states that adopted the shall law in a given year were successful in decreasing three types of crime: violent crimes, robbery and murder. In order to assess this hypothesis, we suggest that the regression model that best captures this causality is of the following form:

$$\boldsymbol{Ln(crime\_rate) = \alpha + \delta\, shall_{s,t} + \sum_{K\,=\,Alabama}^{Wyoming} \beta_s\, stateid_{k,s} + \sum_{j\,=\,1977}^{1999} \gamma_t \,.\, year_{j,t} + \eta\, Controls + e_{d,t}}$$

- $\boldsymbol{stateid_{k,s} = 1\ only\ if\ obs\ is\ from\ stateid\ k, otherwise\ if\ s \neq k\ we\ assign\ 0}$.
- $\boldsymbol{year_{j,t} = 1\ only\ if\ obs\ is\ from\ year\ "j", otherwise\ if\ t \neq j\ we\ assign\ 0}$

We consider the previous model the best approach since allows to capture the effect of issuing the shall-issue laws ($\boldsymbol{\delta}$) in a given state ***(s)*** at a given year ***(t)***, while controlling for those unobservable state differences that are time-invariant by introducing a set of dummy variables for each state ($\boldsymbol{stateid_{k,s}}$). Also, since not all states adopted the law at the same time, we allow for unobservables differences among years that are state-invariant by introducing a set of dummies variables for the year ($\boldsymbol{year_{j,t}}$). We also include a set of controls that are available per state and year in order to explain those differences that are measurable. As an example, the estimated regression for the state of Alabama in the year 1977 will look like:

$$\boldsymbol{Ln(\widehat{crime\_rate}) = \hat{\alpha} + \hat{\delta}\, shall_{AL,77} + \widehat{\beta_{AL}} + \widehat{\gamma_{77}} + \hat{\eta}\, Controls}.$$

In order to assess the previous model, we rely in panel data methods and compare this approach to other types of regressions. We want to test our hypothesis that shall laws reduce crime; then, we define the null ($\boldsymbol{Ho\colon \delta = 0}$) versus the alternative ($\boldsymbol{Ha\colon \delta \neq 0}$) and want to reject the null at least at the 5% level, meaning that the coefficient of shall is statistically significant at the 5% level. When using panel methods, the standard errors should be clustered at the entity level, in section results we discuss this in detail.

Data Section.

We use a cross-sectional time series dataset (panel data) from the year 1977 to 1999 for the states in the USA plus the district of Columbia. We considered three types of crime: violent crimes (*vio*), robbery (*rob*) and murder (*mur*); all of them are given in rates per 100,000 inhabitants. These crime rates are considered as the response variable and we want to find a causal relationship between each of them with the principal explanatory variable *shall;* this is a dummy variable, it takes the value of 1 for those states that adopted the shall-issue laws in a given year, and 0 otherwise. Since we are going to construct different regressions in order to assess the impact of shall issue laws in these crimes, we rely on a set of covariates that are used as controls.

These controls include the variables *incarc_rate* which is the number of prisoners sentenced per 100,000 people lagged one year. A set of demographic variables (*pm1029, pw1064, pb1064*) measuring the percent of males in the population of a given state (*pm*) in ages of 10 to 29, and the percent of the state population that is: white (*pw*) and black (*pb*) in the range of 10 to 64 years of age. *Density* measures the density of population for a given state and year by mil^2 in thousands of people, equally *pop* considers the state population for a given year in millions of people. *Avginc* is measured in 1000s of USD$ for the real per capita personal income at the given state and year. The last two variables, *year* and *stateid*, respectively are ordinals variables for the years (1977 – 1999) and states in the USA in alphabetical order (AL = 1, AK = 2, etc).

Summary Statistics.

We can see that violent crimes (*vio*) are the most committed crime in the USA for the years of 1977 to 1999. The average of this crime (503.07) overcomes the average of robbery (161.82) and murder (7.66) by 67.84% and 98.47%, respectively, making this the most committed crime in the USA, on average. Also, the rates of robbery and murder are more widely dispersed than the rates of violent crimes. For violent crimes one standard deviation is only equivalent to 60% of its mean while robbery and murder each have a standard deviation approximately equivalent to the mean of each. For this sample, these rates are not shared smoothly between the states, there are states with crime rates as high as 2921.0 and as low as 47.0, for other types of crime we have the same problem: robbery (1635, 6.4) and murder (80.6, 0.2). Therefore, to control for possible outliers that may affect our regression results we smooth these rates by applying a natural log transformation on these three crime rates. The resulting transformation apparently does a good job stabilizing the variance and the logged rates now are more appropriate to our inference aims. The same reasoning is applied when transforming the variable *avginc,* this variable contains high outliers, so we take the natural log of it to make the distribution of average income more normal-like. This is usually the case since some states tend to be richer than others, so it makes difficult to compare them. In this sample the average percent of white population (62.94) in the ages of 10 to 64 is almost 91% higher than the average percent of black people (5.34) in the same age range in this sample. The distribution of the average percent of black population is less disperse (1sd = 4.87) than the counterpart white population (1sd = 9.76). Interestingly, there are states that tend to have high incarceration rates (max in the sample = 1913) while others have a rate of 20 points. The problem of the exogeneity of this variable is discussed below. Finally, the average density population in the USA per thousand people between 1977 to 1999 was 0.352 per mile square with a total population average of 4.81 million people.

Section Results

We begin our analysis in table 1 by a naïve comparison, we run a simple linear regression, using OLS method, of the natural logarithm of violent crimes rates on the shall variable. This regression tells us that those states that adopted the shall-issue laws tend to reduce the rates of violent crime by 44.29% (or exact coefficient of 35.79%) points. The coefficient is significant at even less than the 0.1% level but the R^2 is relatively low, this regression while simple and significant only explains 8.66% of the variation in the natural log of violent rates. This means that there are failures in the casual condition and maybe the error term is absorbing all of this variation, i.e., omitted variable bias. Therefore, we run the regression stated in model 2 using the set of controls mentioned before, all of the coefficients obtained are significant at less than the 0.1% level except for *density, avginc* and *pm1029*. Once again, the coefficient in shall is negative and significant but decreased in magnitude so this means that the omission of the controls tends to bias negatively the coefficient in shall (i.e., overestimating the effect of shall issue laws in the decreasing of violent crime rates). While the last regression seems to fit well the data one cannot assume that all the states in the USA behave in the same manner. There might be states that are more violent than by the mere culture of their people. The difficulties now are that one cannot simply measure these differences. The most appropriate way to measure the differences between states is to assume that the unobserved error "e" between the states is composed of two parts one explained by the differences across states that are fixed over time called the unobserved state heterogeneity "a_i" and "u_{it}" that measures the unobservables errors in the states that vary over time we refer to them as idiosyncratic error.

If we are going to control for fixed differences among the states, we also need to cluster for standard errors at the state level since the observations in all the variables are taken at the state specific level and the fixed effect model per se is a clustered mechanism, so failure to cluster our standard errors leads to biased coefficients.

Therefore, the following models includes all of those state specific fixed effects and it clusters the standard errors at the state level to control for heteroskedasticity and autocorrelation-consistent errors among the states, since the errors may be correlated. Autocorrelation of errors is more serious when we consider the time fixed effects; so, for now on we are going to cluster at the state level for the remaining regressions. Then, model 3 includes the fixed effect and we conclude that the former model with controls fails to capture abruptly for the real effect of shall-issue laws in the violent crimes rate. In more detail, we say that the failure to consider the fixed time-invariant differences among states bias our coefficient in *shall* negatively. This results in overestimating its effect on violent crime rates by 32.23% points more than in reality it causes when we control for the fixed effects and the set of covariates. Model 4 shows that if we add controls for time fixed effects, the impact of shall issue laws in the violent crime rate becomes less significant, while controlling for other variables. Fail to include time fixed effects bias negatively our coefficient *shall* but this time the negative bias are smaller than before. This might be the case that, in fact, the coefficients in the dummy variables for each year (minus 1977) are not significant making them not relevant for our model, to measure this we run a statistical test. So, we test the null hypothesis that the set of dummies in year are not significant ($\boldsymbol{Ho\colon All\,Time\,Dummies\,=\,0}$) versus the alternative that they in fact are significant to our model ($\boldsymbol{Ha\colon At\,least\,one\,Time\,Dummies\,\neq\,0}$), the result in the heteroskedastic-robust F-statistic implies that they are significant at even less than the 0.001% level telling us that the set of year dummies are in fact significant to our model. If we set to end our analysis at this point, we may conclude that there is not enough statistical evidence to conclude that shall-issue laws help to reduce the violent crimes in the USA in the period between 1977 to 1999. This is what the authors showed in their research, they also conducted another series of regressions to see the actual relation of shall issue laws in the other crimes.

Section Extension

Then we create table 3, following the same set up of model 4 we create model 5-7 where the only change is the inclusion of the variable *Ln_avginc* instead of *avginc.* From the three regressions in table three we appreciate that the coefficient of *shall* is negative for violent crimes and murder while for robbery is positive, however all of them are not significant not even at the 10% level. The positive case corresponds to the dependent variable *ln_(rob),* which is pretty interesting, since this is telling us the states that permitted the inhabitants to carry guns decreased the rates in violent crimes and murder by around 2% but the rates in Robbery increased by 3.25% these results are not logical. We claim if robberies increased due to the shall laws so violent and murder should increase too, since now our interactions between people and thieves include an increased probability of both of them carrying guns. We are suspicious about the inclusion of the demographic variables, this is due to the possible high multicollinearity among these variables, that is *pm1029* already includes white and black males in the range of 10 to 29 years, so seems not plausible to include a separate set of this variable or even include this subset into another variable such as *pw1064* or *pb1064*. So, we conduct two more test one to measure the jointly relevance of the set of demographic variables in our regressions and another to measure the jointly significance of the variables *pw1064* and *pb1064*. The last test is proposed since the authors Ayres and Donohue III state that the 90% of crime is committed by men, and these variables (*pw1064* and *pb1064*) are counting for some proportion of males that are already in our variable *pm1029.* Then, in table 4 we removed the set of demographic factors that are not relevant for each regression. For instance, the regression for robbery should be set in such way where we remove the set of three demographic variables since there are not relevant for this regression, while for violent crimes and murder these three factors are significant together while the "race" variables (*pw1064* and *pb1064*) are not significant not at even the 10% level. Table 4 shows that *Incarc_rate* is another variable that seems to behave strangely among our four tables so far analyzed, the variable is very close to zero and extremely insignificant. The authors suggested that this variable is not exogenous at all; since, those states that experience higher trends of crime, as a result, tend to be the ones who experience higher incarceration rates. Therefore, in table 5 is an extension of model in table 4 with *incarc_rate* excluded. The coefficients of shall in violent crimes and murder remains the same in magnitude with the same negative sign and not significant, the same is true for robbery but with a positive sign. The main difference occurs at the coefficients of density for the three crimes, all of them become significant for *violent crimes* (negative) and *robbery* (positive) at the 5%. The coefficient in *density* for *murder* while less negative than before increases it significance at the 0.1% level. In model 12, the regression for violent crimes, has the variable in pm1029 significant at the 5% level and the overall $\overline{R^2}$ slightly increases from the model 5. Similarly, the regression for murder has now the variable *Ln_avginc* significant at even less than the 0.1% level and pm1029 significant at the 5% level. The results presented in table 5 reflects that including the wrong variables, those that share multicollinearity or simply not significant for our regression purposes, lead to biased coefficients that otherwise may be significant and help to link casual process.

In the three regressions, for table 5, the coefficient for shall remains non-significant even at the 10% level with clustered standard errors, same results as the authors. Then we fail to find a casual relationship in the issuing of shall-carry guns on the three types of crimes here analyzed (violent crimes, robbery and murder) when controlling for state and time fixed effects and the set of covariates given.

Conclusion

Thus, we use the panel data from 1977 to 1999 in order to find any statistically significant relationship between the decrease of crime and the issue of law that allow citizens to carry guns.

Using a set of covariates and panel data methods we run a number of different specifications and after removing those variables not useful for our estimation process, we fail to find any statistically significant causal effect of shall issue laws on the crime rates for the given states between the years 1977 to 1999. The problem for the panel methods in these specifications might be due to the short set of variables and the fact that some of them are already not useful to predict a casual relationship between shall issue laws and crime. Future research on this topic has to include a higher number of relevant controls and an extension of years to include observations near to the present date.

References

Ayres, I. and Donohue III, J.J. (2003). Shooting Down the More Guns, Less Crime Hypothesis. *Stanford Law Review,* 55 (4): 1193–1312. Retrieved from https://digitalcommons.law.yale.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=2240&context=fss_papers.

Lott, J. R. (2010). *More guns, less crime: Understanding crime and gun-control laws*. Chicago: The University of Chicago Press.

Appendix

Table 1: *Summary Statistics*

	mean	**sd**	**max**	**min**
vio	503.075	334.277	2921.8	47.0
rob	161.820	170.510	1635.1	6.4
mur	7.665	7.523	80.6	0.2
shall	0.243	0.429	1.0	0.0
incarc_rate	226.580	178.888	1913.0	19.0
density	0.352	1.355	11.1	0.0
avginc	13.725	2.555	23.6	8.6
pop	4.816	5.252	33.1	0.4
pb1064	5.336	4.886	27.0	0.2
pw1064	62.945	9.762	76.5	21.8
pm1029	16.081	1.732	22.4	12.2
Ln(vio)	6.027	0.646	8.0	3.9
Ln(rob)	4.685	0.955	7.4	1.9
Ln(mur)	1.783	0.703	4.4	-1.6
Ln(avginc)	2.603	0.181	3.2	2.1
N	1173			

Table 2: *Regression Analysis of the log of the violent crime rate and shall-issue laws.*

	(1)	(2)	(3)	(4)
shall	-0.4430*** (0.0475)	-0.3684*** (0.0348)	-0.0461 (0.0418)	-0.0280 (0.0407)
incarc_rate		0.0016*** (0.0002)	-0.0001 (0.0003)	0.0001 (0.0002)
density		0.0267^{+} (0.0143)	-0.1723 (0.1376)	-0.0916 (0.1239)
avginc		0.0012 (0.0073)	-0.0092 (0.0130)	0.0010 (0.0165)
pop		0.0427*** (0.0031)	0.0115 (0.0142)	-0.0048 (0.0152)
pb1064		0.0809*** (0.0200)	0.1043** (0.0327)	0.0292 (0.0495)
pw1064		0.0312** (0.0097)	0.0409** (0.0135)	0.0093 (0.0238)
pm1029		0.0089 (0.0121)	-0.0503* (0.0207)	0.0733 (0.0525)
States Effects	No	No	Yes	Yes
Time Effects	No	No	No	Yes
Clustered s.e.'s	No	No	Yes	Yes
Time Effects = 0 (p-value)				21.6216 0.0000
Demogrphcs = 0 (p-value)		14.1418 0.0000	4.6906 0.0058	3.4211 0.0241
R^2 *adj*	0.0859	0.5613	0.2124	0.4027

These regressions are estimated using the panel data for all the states in the USA plus the district of Columbia. The dependent variable is the natural log of Vio. All regressions use data from 1977 to 1999. Clustered standard errors at the state level. Test on year dummy variables coefficients under the null Time Effects = 0. This panel data can be encountered in Ayres.dta. Note that the standard errors are given in parenthesis under coefficients and p-values are given in parenthesis under the F-statistics. The significance codes used are given at the $+\ p < 0.10$, $*\ p < 0.05$, $**\ p < 0.01$, $***\ p < 0.001$ significance levels.

Table 3: *Regression Analysis for the log of the three types of crime and shall-issue laws.*

	(5) **Vio**	**(6)** **Robbery**	**(7)** **Murder**
shall	-0.0245 (0.0410)	0.0325 (0.0520)	-0.0234 (0.0385)
incarc_rate	0.0001 (0.0002)	0.0000 (0.0003)	-0.0001 (0.0004)
density	-0.0788 (0.1238)	-0.0282 (0.1954)	-0.5956^{+} (0.3248)
pop	-0.0033 (0.0147)	0.0029 (0.0237)	-0.0334 (0.0213)
Ln_avginc	0.2215 (0.2956)	0.6731 (0.4699)	0.8895^{***} (0.2282)
pb1064	0.0330 (0.0484)	0.0208 (0.0811)	0.0145 (0.0769)
pw1064	0.0113 (0.0231)	-0.0101 (0.0323)	-0.0081 (0.0201)
pm1029	0.0654 (0.0529)	0.0874 (0.0715)	0.0687 (0.0410)
States Effects	Yes	Yes	Yes
Time Effects	Yes	Yes	Yes
Clustered s.e.'s	Yes	Yes	Yes
Time Effects = 0 (p-value)	23.6885 (0.0000)	26.6273 (0.0000)	19.2674 (0.0000)
Demographics = 0 (p-value)	3.0679 (0.0362)	0.9600 (0.4190)	2.5872 (0.0633)
Race = 0 (p-value)	0.2343 (0.7920)		0.2799 (0.7570)
R^2 *adj*	0.4055	0.2320	0.2739

These regressions are estimated using the panel data for all the states in the USA plus the district of Columbia. The dependent variables are the natural log of Vio, Rob and Mur respectivelly. All regressions use data from 1977 to 1999. Clustered standard errors at the state level. Test on year dummy variables coefficients under the null Time Effects = 0. Test on pb1064 = pw1064 = pm1029 = 0 under the null Demographics = 0 and pb1064 = pw1064 = 0 under the null Race = 0. This panel data can be encountered in Ayres.dta. Note that the standard errors are given in parenthesis under coefficients and p-values are given in parenthesis under the F-statistics. The significance codes used are given at the
*+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$* significance levels.

Table 4: *Regression Analysis of the log of the three types of crime and shall-issue laws, some variables omitted.*

	(8) **Vio**	**(9)** **Robbery**	**(10)** **Murder**
shall	-0.0227 (0.0428)	0.0531 (0.0566)	-0.0296 (0.0391)
incarc_rate	0.0000 (0.0002)	0.0000 (0.0003)	-0.0001 (0.0004)
density	-0.0678 (0.1312)	0.1167 (0.1924)	-0.5043^{*} (0.2191)
pop	-0.0059 (0.0112)	0.0152 (0.0182)	-0.0269 (0.0213)
Ln_avginc	0.2044 (0.2840)	0.8362^{+} (0.4507)	0.9109^{***} (0.2352)
pm1029	0.0816^{*} (0.0333)		0.0597^{*} (0.0269)
States Effects	Yes	Yes	Yes
Time Effects	Yes	Yes	Yes
Clustered s.e.'s	Yes	Yes	Yes
Time Effects = 0 (p-value)	45.0538 (0.0000)	36.2208 (0.0000)	21.1041 (0.0000)
R^2 *adj*	0.4053	0.2125	0.2730

These regressions are estimated using the panel data for all the states in the USA plus the district of Columbia. The dependent variables are the natural log of Vio, Rob and Mur, respectivelly. All regressions use data from 1977 to 1999. This panel data can be encountered in Ayres.dta. Clustered standard errors at the state level. Note that the standard errors are given in parenthesis under coefficients and p-values are given in parenthesis under the F-statistics. The significance codes used are given at the $+\ p < 0.10$, $*\ p < 0.05$, $**\ p < 0.01$, $***\ p < 0.001$ significance levels.

Table 5: *Regression Analysis of the log of the three types of crime and shall-issue laws, incarc_rate omitted.*

	(11) **Vio**	**(12)** **Robbery**	**(13)** **Murder**
shall	-0.0228	0.0529	-0.0294
	(0.0429)	(0.0563)	(0.0389)
density	-0.0945^{*}	0.0914^{*}	-0.4596^{***}
	(0.0395)	(0.0421)	(0.0424)
pop	-0.0049	0.0162	-0.0287
	(0.0118)	(0.0166)	(0.0207)
Ln_avginc	0.2053	0.8364^{+}	0.9093^{***}
	(0.2840)	(0.4512)	(0.2356)
pm1029	0.0812^{*}		0.0605^{*}
	(0.0333)		(0.0262)
States Effects	Yes	Yes	Yes
Time Effects	Yes	Yes	Yes
Clustered s.e.'s	Yes	Yes	Yes
Time Effects = 0	45.2508	36.7811	21.3709
(p-value)	(0.0000)	(0.0000)	(0.0000)
R^2 *adj*	0.4056	0.2131	0.2734

These regressions are estimated using the panel data for all the states in the USA plus the district of Columbia. The dependent variables are the natural log of Vio, Rob and Mur, respectivelly. All regressions use data from 1977 to 1999. Clustered standard errors at the state level. Test on year dummy variables coefficients under the null Time Effects = 0. Test on pb1064 = pw1064 = pm1029 = 0 under the null Demographics = 0 and pb1064 = pw1064 = 0 under the null Race = 0. This panel data can be encountered in Ayres.dta. Note that the standard errors are given in parenthesis under coefficients and p-values are given in parenthesis under the F-statistics. The significance codes used are given at the
*+ p < 0.10, * p < 0.05, ** p < 0.01, *** p < 0.001* significance levels.